Localization in ensemble data assimilation

P. Kirchgessner[1], L. Nerger[1] and A. Bunse-Gerstner[2]

[1]Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
[2]University of Bremen, Germany

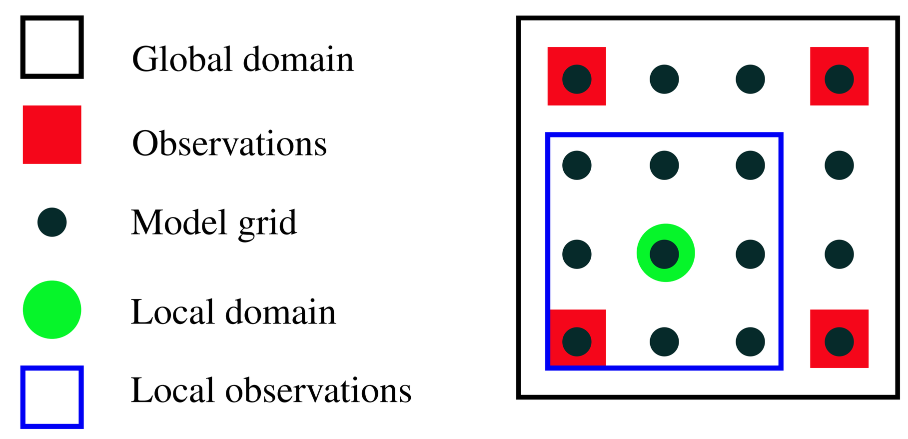
Introduction

In data assimilation using ensemble Kalman filter methods, localization is an important technique to get good assimilation results. For the LETKF [1], the domain localization (DL) and observation localization (OL) are typically used. Depending on the localization method, one has to choose appropriate values for the localization parameters, such as the localization length, the inflation factor or the weight function. Although being frequently used, the properties of the localization techniques are not fully investigated. Thus, up to now an optimal choice for these parameters is a priori unknown and they are generally found by doing expensive numerical experiments.
The relationship between the localization length and the ensemble size in DL and OL is studied using twin experiments with the Lorenz-96 model [3]. It is found that for DL the optimal localization length depends linearly on the local observation dimension. This also holds for the localization length at which the filter diverges. A similar behavior was observed for OL by considering an effective local observation dimension.

Localization Techniques for LETKF

Domain localization (DL)

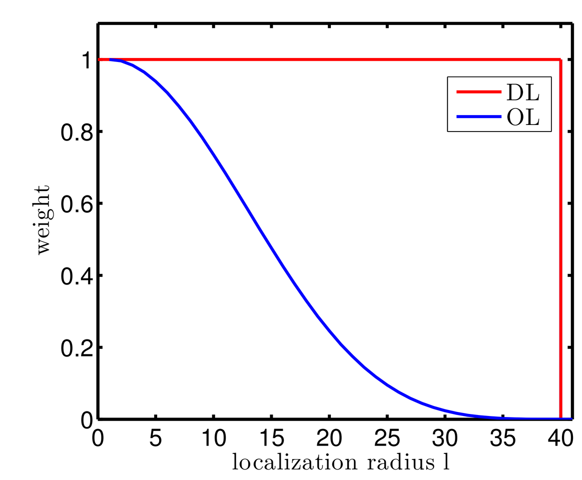
- Decompose the whole domain G in smaller domains G_i.
- Choose a domain D_i in observation space within the observations are relevant for the analysis in G_i.
- For all different domains G_i calculate an analysis with the observations in D_i.
- Restore the global state for the next forecast.



Observation Localization (OL)

- Decompose the whole domain G in smaller domains G_i.
- Weight the observations depending on the distance to the analysis point (e.g. with the 5-th order polynomial [3]).
- For all different domains G_i calculate an analysis with the observations in D_i.
- Restore the global state for the next forecast.

Right: The weight functions used for DL (red) and for OL (blue).



Experimental Setup

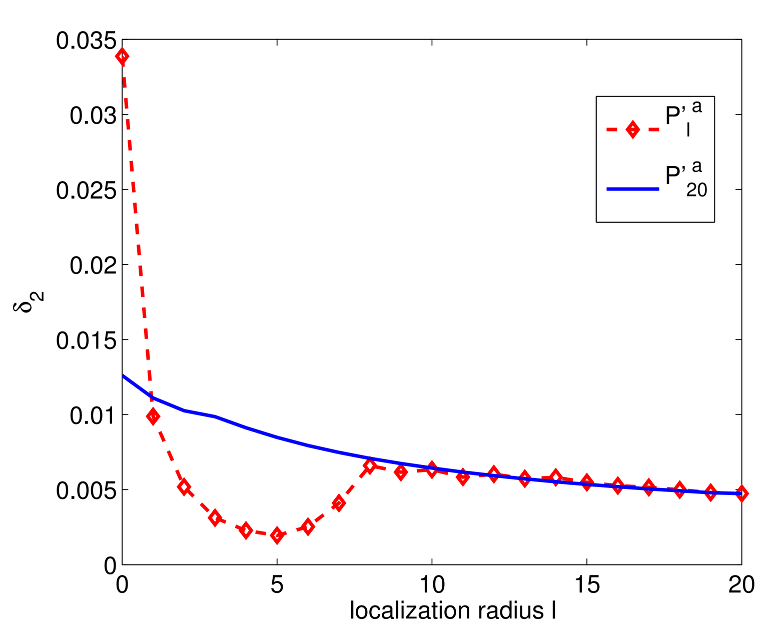
Filter Configuration

Assimilations were performed by using the LETKF [1] with DL and OL. In each step the whole state was observed. The ensemble was generated by choosing random states from a long model run. The domain decomposition was made by calculating a separate analysis for every single state component. Observations within the localization radius l were used for the assimilation each model grid point. The localization radius l was varied from 1 to 20 and the number of ensemble members from 5 to 30. For OL, the observations were weighted by using the fifth order polynomial introduced by Gaspari and Cohn [3], for several localization radii.

Description of experiments

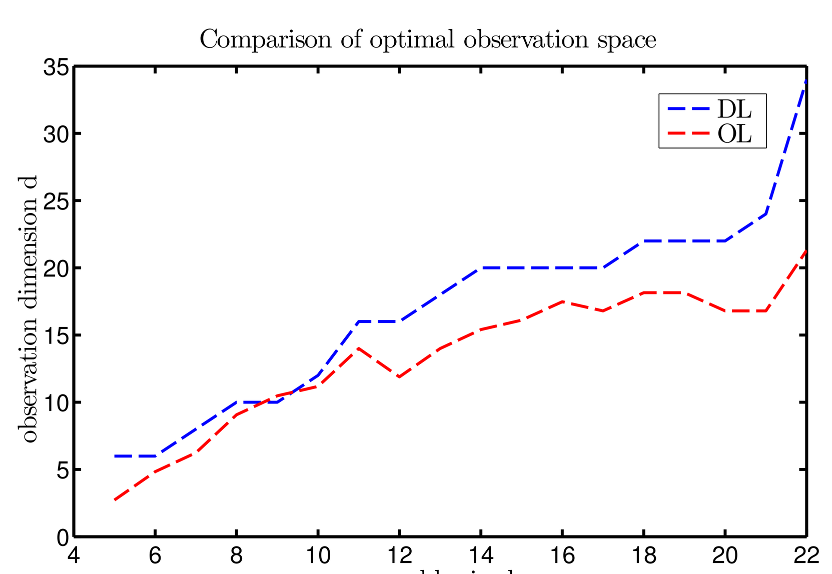
Twin experiments for various sets of parameters for OL have been performed. The observations, generated with a standard deviation $\sigma_o = 1$, have been assimilated for 5000 consecutive time steps. For statistical significance, all experiments were repeated $10-20$ times. The experiments have been performed with PDAF [4].
The results have been evaluated by calculating the mean RMS error of the analysis estimates

Sampling quality



Left The improved analysis correlates with an improved estimate of the covariance matrix. This was observed by considering the difference δ_2 between an ideal covariance matrix and the estimate. If the localization radius is to small, the analysis is improved, but the covariance is not well estimated. For moderate localization radii the covariances are better estimated, therefore the analysis becomes better.

Conclusion



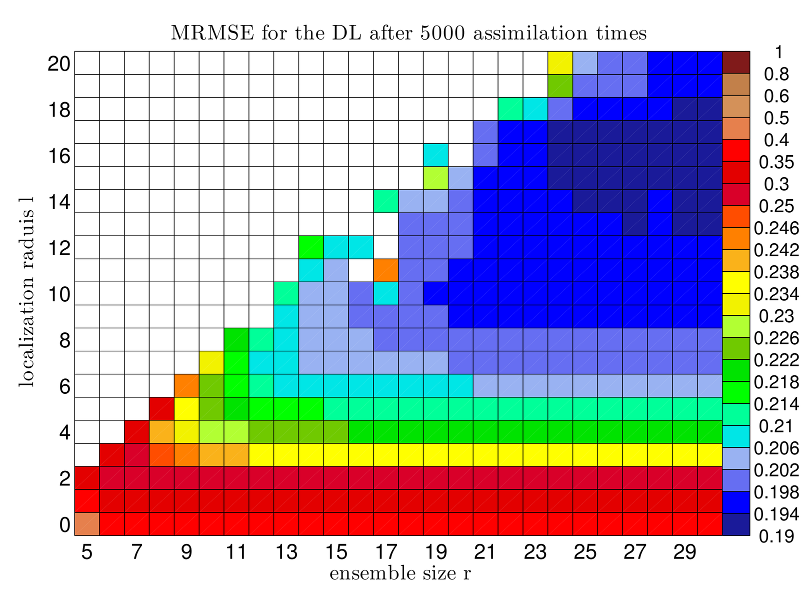
Left By considering the sum of the weights of the weighting function as an approximation to the observation dimension, it is possible to relate the results for both localization techniques. For both methods the curves show similar behavior. This explains the difference in observed behavior between the two methods.

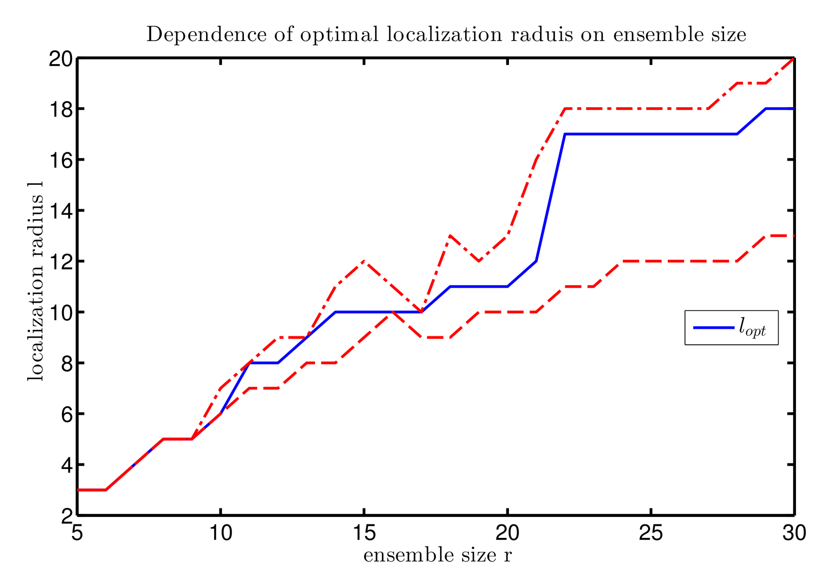
Results

Domain localization

Below Each field in the matrix stands for the mean RMS error (MRMSE) of a certain configuration. A white entry means filter divergence. In most cases filter divergence happens if the number of observations exceeds the number of ensemble members.
The gain by increasing the ensemble size is very limited if the localization radius l is kept constant. More improvement can be achieved by choosing the optimal localization radius.

The optimal localization radius is nearly linear dependent on the number of ensemble members. The region where the difference is less than 1% from the optimal configuration widens for increasing ensemble size. In the case where the localization radius is much smaller then the ensemble size, the optimal interval is very narrow and the localization radius has to be carefully chosen in order to get optimal results.

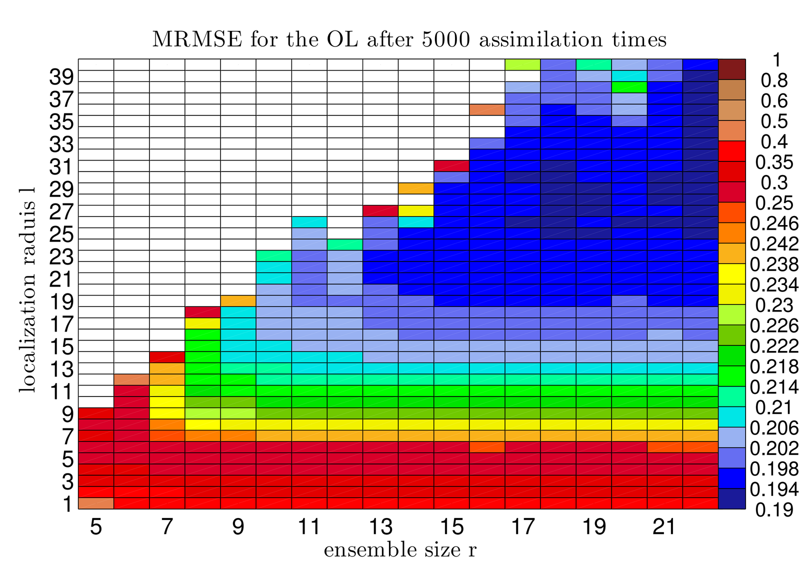


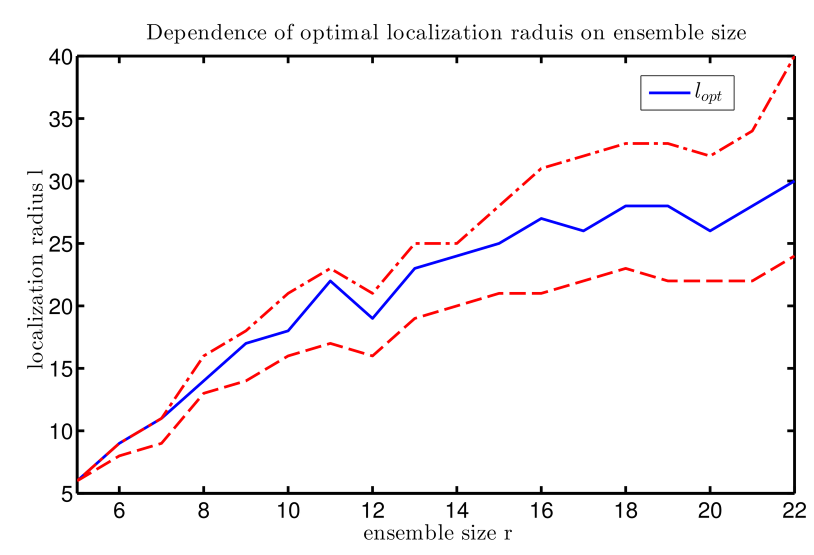


Observation localization

Below The relationship between the ensemble size r and the localization radius l for OL is similar to DL. If the localization radius is increased too much, the filter diverges. In contrast to DL, l can be chosen bigger before this happens.

If the localization radius l is small compared to the ensemble size r there is a linear dependence of l_{opt} and r. As for OL, the optimal interval widens for increasing l. Compared to OL, l_{opt} can be chosen slightly bigger.





References

[1] Hunt, B.R., E.J. Kostelich, and I. Szunyogh (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* 230: 112–126

[2] G. Gaspari, S. E. Cohn (1999). Construction of correlation functions in two and three dimensions *Q. M. R.* DOI: 10.1002/qj.49712555417

[3] . E.N. Lorenz (1996). Predictability: a problem partly solved *In: Proceedings of the Seminar on Predictability* ECMWF , Reading , UK, 1-18

[4]L. Nerger and W.Hiller (2012). Software for Ensemble-based Data Assimilation Systems. Computers and Geosciences. In press. doi:10.1016/j.cageo.2012.03.026